



RDM FOR DATA FROM HPC

15.04.2024 | Sander Apweiler, Benedikt von St. Vieth, Rajveer Saini | JSC

1. HPC Systems
2. Stagesystems
3. Transfertools
4. B2SHARE
5. DataPub
6. Jülich DATA
7. Comparison of Data services

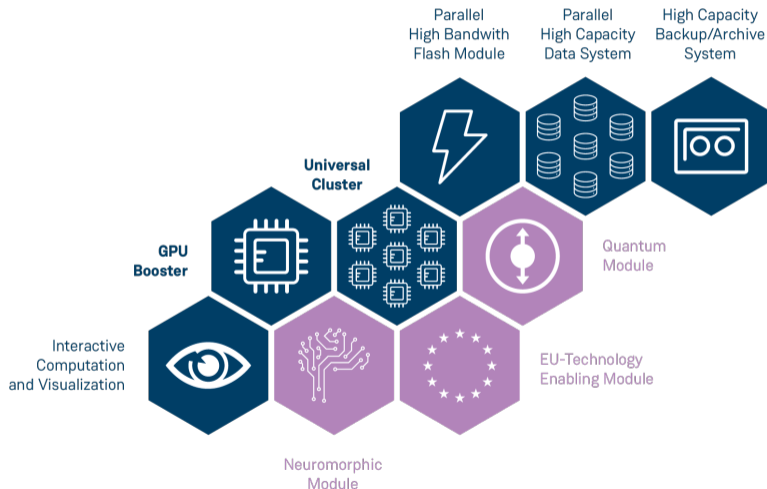
HPC Systems

HPC Systems - ... to the present

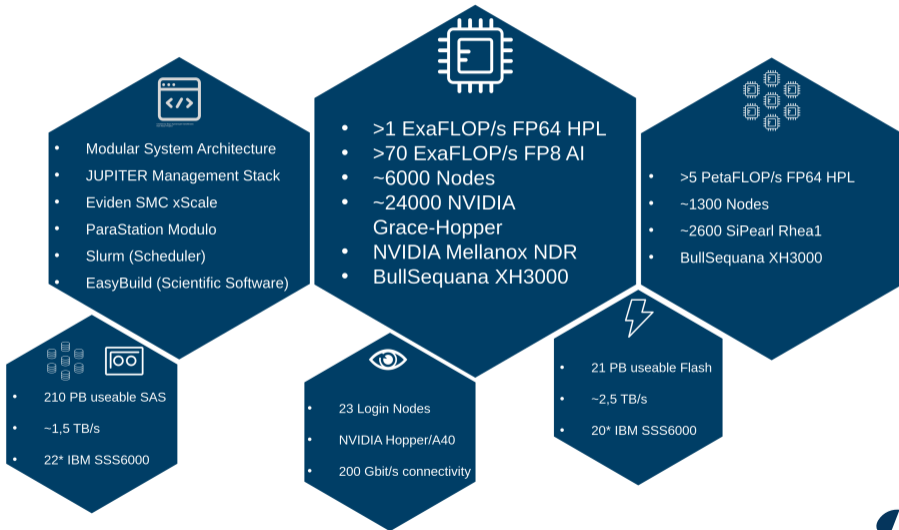
	JUWELS		JURECA	
	Cluster	Booster	Data Centric	AI
Peak Petaflops/s	12,3	75	18,5	1
Nodes	2567	936	768	16
Accelerators	224 GPUs	3744 GPUs	768 GPUs	64 GPUs

- Beside of the large HPC clusters
 - Prototype as a blueprint for modular super computers: DEEP
 - Small cluster for user from campus or specific communities: HDF-ML, JUZEA1
 - Data access cluster: JUDAC
 - Others: JSC Cloud (combination of JUSUF and HDF Cloud)

HPC Systems - Jupiter High Level Architecture



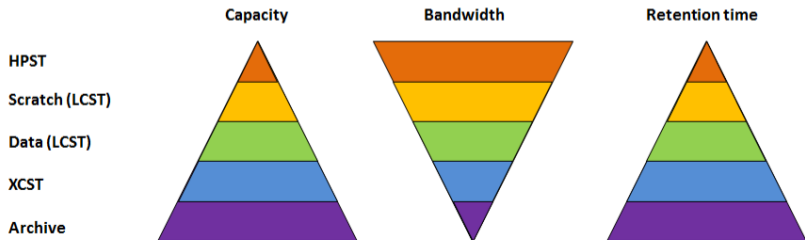
HPC Systems - Jupiter High Level Architecture



Storagesystems

Storagesystems

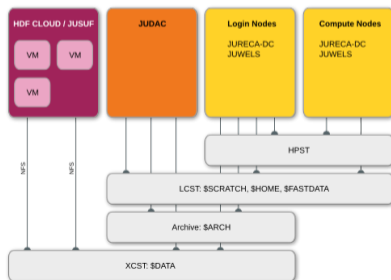
- High Performance Storage Tier (HPST):
NVMe based Storage, 110 servers, 1100 drives
- Large Capacity Storage Tier (LCST):
5th Gen. of JUST, 22 building blocks, 7500 x 10TB discs
- Extended Capacity Storage Tier (XCST):
7 GPFS Building Blocks using 7000 discs (10TB, 12TB, 16TB)
- Archive: Tape storage (Backup + GPFS&TSM-HSM),
1 building block + 3 tape libraries, \approx 34500 tapes



Storagesystems

Filesystems:

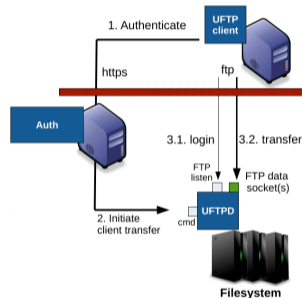
- \$HOME: user's home directory inside GPFS
- \$SCRATCH: temporary storage location for applications with large size and I/O demands, deleted after 90 days
- \$DATA: large capacity for storing and sharing data
- \$FASTDATA: storage for large projects in collaboration with JSC
- \$ARCH: storage for all files not in use for a longer time



Transfertools

Transfertools - UFTP

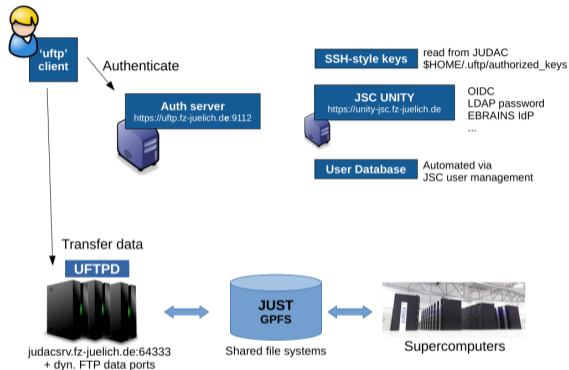
- „UNICORE FTP“ \Rightarrow it grew out of the need for a high-performance data transfer for the UNICORE federated middleware
- Client/server file transfer system, based on FTP, with flexible, federated authentication
- Extended with
 - Compression and/or encryption of data stream(s)
 - Multiple TCP streams per data connection
 - Bandwidth management (both server and client)
- UFTPD Server supports partial file reads/writes
- UFTPD creates multiple sessions in parallel for scalability
- Multiple UFTPD backends are possible (round-robin)



Transfertools - UFTP

- Share data with other users via UFTP
- Command line tool with well known commands like:
cp, ls, ...
- Authentication via
 - ssh-style keys (recommended, at JSC it must differ from ssh key)
 - OAuth tokens
 - username/password
- Available at JSC, HLRS and LRZ

Full presentation about UFTP at
DOI:10.23728/b2share.83b63c4bf4674e638711ef9bddb93bfb or
PID: 11304/7d62b0f3-02a3-4d82-aa47-436fa9778cda



B2SHARE

B2SHARE

- Developed by EUDAT CDI
- Provided by JSC as community instance
- URL: <https://b2share.fz-juelich.de>
- Audience: researchers from Forschungszentrum Jülich and related communities
- Quota: 50GB per file; 200GB per deposit
- Listed as repository on re3data
- Accepted repository for Springer publication
- Harvested by B2FIND

GO TO EUDAT WEBSITE

Search records for...

HELP COMMUNITIES UPLOAD CONTACT

Store and publish your research data

Search in public datasets or register as a user to upload and publish your data!

Create record

Latest records

Representational Models for 3D neuroanatomic segmentation
29 Mar 2022 by Ray, Sakar, Engler, David, Rivier, Martin
With this work, we provide a 3D whole brain segmentation model that augments the 2.5D approach FastSurferCNN. Furthermore, we explore the efficacy of spatial embedding and its parameters. [View](#)

Global, high-resolution mapping of tropospheric ozone - explainable machine learning and impact of uncertainties - Source Code
24 Feb 2022 by Tomer, Chen, Shmida, Yoni, Shiri, Joni, Karim, Parvati, Inat, Shalim, Iva, et al
This source code contains all methods that is being used in ozone-mapping project. In addition, it contains scripts to run both explainable AI methods and methods used to study the impact of uncertainty. [View](#)

Global Aerosol GPP profiles reanalysis from MERRA-2, v6.1
21 Jan 2022 by Gertsen, Fabian, Buchard, Jorgos, De Siro, Helena, Hortal, Pierre, Müller, Marc
Global distribution of monthly aerosol optical properties individual, single scattering albedo and asymmetry parameters created from the latest version 2 primary reanalysis profiles of aerosol mixing. [View](#)

TOAR-II data gathering event, 21 Jan 2022
21 Mar 2022 by Schütz, Martin R., Weisbach, A., Schröder, S., Salts, N., Homburg, M.
The Tropospheric Ozone Assessment Report (TOAR) is an initiative of the international Global Atmospheric Chemistry (GAC) project. TOAR-II is the second phase of TOAR, it builds on the successful TOAR I.

Global Aerosol GPP profiles reanalysis from MERRA-2, v6.1
8 Feb 2022
See Global_Aerosol_GPP_profiles_reanalysis_from_MERRA-2_v6.1 for full description.

Global Aerosol GPP profiles reanalysis from MERRA-2, v6.1
8 Feb 2022
See Global_Aerosol_GPP_profiles_reanalysis_from_MERRA-2_v6.1 for full description.

Global Aerosol GPP profiles reanalysis from MERRA-2, v6.1
21 Jan 2022
See Global_Aerosol_GPP_profiles_reanalysis_from_MERRA-2_v6.1 for full description.

Archival material from the TOAR-II Science Working Group, New 16-18, 2021
30 Nov 2021 by Cooper, Owen R., Schult, Martin G., Ziemann, Uwe, Kuehn, Roger, Shalim, Roman, Shalim, Yoni, von Klenzendorfer, Ulrik, Legent, Bridget, Dörfler, Robert, Iversen, Anders, Zhang, Lin
The Tropospheric Ozone Assessment Report (TOAR) is an initiative of the international Global Atmospheric Chemistry (GAC) project. TOAR-II is the second phase of TOAR, it builds on the successful TOAR I.

GO TO EUDAT WEBSITE

Search records for...

HELP COMMUNITIES UPLOAD CONTACT

RECORDS

TSID Select the schema and enter the deposit title

Community

AUTO
A?

BEAM

CLARIN

DIRAC

EGAT

EUDAT

EURO-CORDEX

EBF

Indivisibility

LTER

NSF

RFA

TOAR

Create Draft Record

You can also update the data in an existing record by creating a new version of that record. Search for the 'Create new version' button on the record's page.

GO TO EUDAT WEBSITE

Search records for...

HELP COMMUNITIES UPLOAD CONTACT

RECORDS - 07284988020001102 BARNEY - 001

Editing draft version

Add files

Drop files here, or click to select files

Add B2DROP files

Basic fields

Community *

TSID * Select the schema and enter the deposit title

Description

Type *

Creators

Open Access * True

Embargo Date

License

URL

Keywords

Contact Email

Publication Date

Show more details

Submit draft for publication
When this draft is published it will be assigned a PID and a DOI, making it publicly citable. Please note that the published record's file can no longer be modified by its owner.
The publication will get the following DOI: [10.34730/07284988020001102.0481917.0](https://doi.org/10.34730/07284988020001102.0481917.0)

This draft is up to date

B2SHARE

- Attaches DOI to each deposit and PID to each deposit and all files
- Supports multiple versions of a deposit
- Supports individual (community) metadata schemas
- Usage with web UI or REST API
- Command line tool for basic usage available
- Link to other data storage is possible, if the data is stored somewhere else
- More information: <https://go.fzj.de/B2SHARE>

[SEARCH](#)[HELP](#) [COMMUNITIES](#) [UPLOAD](#) [CONTACT](#)

sa.apweiler@fz-juelich.de

[RECORDS](#) » 3F5C3F8E853148D2806867A93EECD4E0

Latest Version - Apr 29, 2021

TOAR-II Database Infrastructure presentation for WMO virtual conference on Earth system data exchange in the 21st century, 16 - 19 Nov 2020

by Schroder, Sabine; Epp, Eleonora; Leufen, Lukas; Mozaffari, Amirpasha; Romberg, Mathilde; Schultz, Martin; Sun, Jianing;

Apr 29, 2021

Abstract: Innovative Air Quality Services: The TOAR Database Infrastructure

The Tropospheric Ozone Assessment Report database of global air quality observations includes multi-faceted metadata and was built according to Open Data and FAIR principles.

We now collect OpenAQ data and exploit high-resolution geospatial data.

This presentation was shown at the poster session of the WMO virtual conference on Earth system data exchange in the 21st century, 16 - 19 Nov 2020 (<https://meetings.wmo.int/WMO-Data-Conference/SitePages/Conference%20Scheme.aspx>).**Disciplines:** 3.2.4 → Chemistry → Atmospheric chemistry;**Keywords:** tropospheric ozone, global air quality, air quality monitoring;;**DOI:** [10.34730/3f5c3f8e853148d2806867a93eeed4e0](https://doi.org/10.34730/3f5c3f8e853148d2806867a93eeed4e0)**PID:** [21.11125/dac26910-1581-46c0-9c33-a4131865e827](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-7)

Files

Name

15.04.2024
Size

Basic metadata

Open Access

Slide 15
True ✓

DataPub

DataPub

- URL: <https://datapub.fz-juelich.de/>
- Solution for files which do not fit in B2SHARE
- Audience: researchers from Forschungszentrum Jülich
- Listed as repository on re3data
- Stores only data but no metadata and does not assign identifiers
- Accessible via https
- Datarepository can be managed by single users or groups
- Usage must be requested

Jülich DATA

Jülich DATA

- URL: <https://data.fz-juelich.de/>
- Based on Dataverse
- Service provided by ZB (central library) as central solution
- Connected via OpenID Connect to Helmholtz AAI
- Provides DOIs
- Storage for metadata and small files (up to 2GB)
- Link to other storages with URIs
- Storage tiering if institutes provide their own S3 storage
- Individual metadata schemas are possible

Comparison of Data services

	B2SHARE	DataPub	Jülich Data
Filesize	50GB	unlimited	2GB
Recordsize	200GB	unlimited	-
Metadata	Yes	No	Yes
PID/DOI	Yes	No	Yes
Access	FZJ Internal	External	FZJ Internal
Retention	10Y	10Y	10Y
Registry	re3data, Fairsharing, DataCite	re3data	re3data
External repository	No	Yes	Yes